TRANSPOSON INSERTION POLYMORPHISM AS A NEW SOURCE OF MOLECULAR MARKERS

Dariusz Grzebelus

Dept. of Genetics, Plant Breeding and Seed Science
Agricultural University of Cracow
al. 29 Listopada 54, 31-425 Cracow, POLAND
e-mail: dgrzebel@bratek.ogr.ar.krakow.pl

(Received September 15, 2005/Accepted October 30, 2005)

ABSTRACT

Transposons constitute a significant fraction of plant genomes and are considered to be one of the major forces driving genome evolution. Each transposition event generates new variability. The presence or absence of a transposon at a particular location in the genome can be readily determined by several molecular techniques. Polymorphic insertion sites can be used in the same way as other widely used molecular marker systems. Molecular marker systems based on transposons can be used to investigate diversity and to create genetic linkage maps. New families of transposons are being identified and characterized. They can be used to develop new marker systems. The present paper reviews approaches which have been used to identify polymorphic transposon insertion sites.

Key words: transposable elements, transposon display, diversity, genetic linkage mapping

INTRODUCTION

Transposons are mobile genetic elements capable of changing their location in the genome. They were discovered almost sixty years ago in maize (McClintock, 1948). However, only in the last twenty or thirty years has interest in transposons grown now that the genomes of several model organisms have been sequenced. Our perception of transposons has changed dramatically from "selfish DNA, the ultimate parasite" (Orgel and Crick, 1980) to "important players in the evolution of plant genomes" (Wessler et al., 1995).

There are two general classes of transposons:

Class I consists of retrotransposons, which transpose in a replicative manner using an RNA intermediate. This means that each transposition event

creates a new copy of the transposon while the original copy remains intact at the donor site.

Class II consists of DNA transposons, which change their location in the genome by a 'cut and paste' mechanism. This means that they excise themselves from the donor site and reintegrate themselves at the acceptor site.

Based on structural characteristics, transposons can be further subdivided into subclasses, superfamilies, families, and subfamilies. The most important criteria are:

1) the type and orientation of open reading frames carried by the transposon and encoding for proteins required for transposition, such as reverse transcriptase and integrase in Class I transposons, and transposase in Class II transposons;
2) the presence, orientation, length and sequence of their terminal repeats; and
3) the length and sequence of TSDs (Target Site Duplications) created upon insertion.

The main retrotransposon superfamilies are *Ty1/copia*-like, *Ty3/gypsy*-like, *LINE* (*Long Interspersed Nuclear Elements*) and *SINE* (*Short interspersed Nuclear Elements*). There are several well-studied families of DNA transposons in plants, including the *Ac/Ds*, *Mutator*, *En/Spm* and *PIF* families in maize.

Each transposon family usually consists of a few autonomous transposons, all of which carry structural components needed for transposition. Each family also usually contains some non-autonomous transposons with internal deletions which are capable of transposition only after cross-mobilization by the related autonomous transposon (Capy et al., 1998).

In addition, there are also some transposons which cannot be assigned to Class I or Class II. One large group of unclassifiable transposons are *MITEs* (*Miniature Inverted-Repeat Transposable Elements*), which are small, not exceeding several hundred base pairs. *MITEs* are usually present in the genome in large numbers, sometimes in more than ten thousand copies. They were recently shown to have several structural characteristics in common with certain Class II transposons. This suggests that they probably share a common ancestor. This hypothesis is further supported by the fact that, in rice, a *MITE*-like transposon *mPing* was cross-mobilized by a distantly related Class II transposon *Pong* (Jiang et al., 2003; Kikuchi et al., 2003). However, there is no good explanation how *MITEs* attained high copy numbers considering that "cut and paste" transposition usually does not generate new copies of transposons.

Each transposition event generates an insertion polymorphism which can be identified using a range of molecular techniques. Transposons present in numerous copies can generate hundreds, if not thousands, of markers. The presence or absence of the transposon at a given locus can be used as

a molecular marker for genotype fingerprinting, investigations on diversity and phylogeny, and genetic linkage mapping. If an active transposon has been identified or introduced into the genome by transformation, these markers can also be used to target novel insertion sites giving rise to a mutant phenotype.

The present paper briefly reviews different approaches used to generate molecular markers based on transposon insertion polymorphism.

Systems used for the identification of transposon insertion sites

Several systems for identifying transposon insertion sites have been developed over the last ten years. Although they differ from a technical point of view, in principle, they all provide the same type of information. They can be used to identify transposon insertion sites using PCR amplification or hybridization. Brief descriptions of all the systems which have been reported in the literature are presented below.

Inter-Retrotransposon Polymorphism (IRAP) and Retrotransposon-Microsatellite Polymorphism (REMAP)

Both IRAP and REMAP were first described by Kalendar et al. (1999). IRAP is based on the PCR amplification of genomic DNA fragments which lie between two retrotransposon insertion sites. REMAP is based on the amplification of fragments which lie between a retrotransposon insertion site and a microsatellite site. In both techniques, polymorphism is detected by the presence or absence of the PCR product. Lack of amplification indicates the absence of the retrotransposon at the particular locus.

Transposons with high copy numbers are the best transposons for IRAP and REMAP. The technique was originally developed using the *BARE-1* retrotransposon, which is present in the barley genome in numerous copies. About thirty bands were visualized following a single PCR reaction (Kalendar et al., 1999). These markers were extremely polymorphic, which made them useful for evaluating intraspecific relationships. The same system was successfully used to map a major resistance gene for net blotch in barley (Manninen et al., 2000).

Inter-MITE Polymorphism (IMP)

IMP is in principle very similar to IRAP, except that it uses *MITE*-like transposons rather than retrotransposons. It can generate very complex electrophoretic profiles. It was first used to identify two groups of *MITEs* in barley, one belonging to the *Stowaway* family, and the other to the recently identified *Barfly* family (Chang et al., 2001). The results were used to assess genetic similarity among several barley cultivars, and to fill out the linkage map of the barley genome.

Sequence-Specific Amplification Polymorphism (S-SAP)

S-SAP was first used to investigate the location of *BARE-1* retrotransposons in the barley genome (Waugh et al., 1997). In principle, it is a simple modification of the standard AFLP (Amplified Fragment Length Polymorphism) protocol (Vos et al., 1995). The final amplification is performed with retrotransposon-specific and *Mse*I-adaptor-specific primers. S-SAP has been extensively used to generate markers to study genetic diversity and to prepare linkage maps in several plants, including the pea, *Medicago,* wheat, and the cashew (Pearce et al., 2000; Porceddu et al., 2002; Queen et al., 2004; Syed et al., 2005). All S-SAP systems described to date utilize LTRs (Long Terminal Repeats) of *Ty1-copia* or *Ty3-gypsy* retrotransposons. They were all proved to be as efficient as or even more efficient than the original AFLP technique.

IRAP, REMAP and S-SAP can all be used for mapping, fingerprinting, marker-assisted selection and evolutionary studies (Leigh et al., 2003). It is very important to select the right retrotransposon family, because different families may differ in the amount of information they provide and the quality of the electrophoretic profiles they produce.

Retrotransposon-Based Insertion Polymorphism (RBIP)

RBIP differs from the methods described above because it requires the sequence of the 5' and 3' regions flanking the transposon insertions to be completely known. When a primer specific to the transposon is used together with a primer designed to anneal to the flanking region, they generate a product from template DNA containing the insertion. On the other hand, primers specific to both flanking regions amplify a product if the insertion is absent. Polymorphisms can be identified using standard agarose gel electrophoresis, or by hybridization with a reference PCR fragment. Hybridization is more useful for automated, high throughput analysis.

RBIP was developed using the *PDR1* retrotransposon in the pea (Flavell et al., 1998). It is much more costly and technically complicated than other methods for detecting transposon insertions. However, it generates co-dominant markers, which can be at least as useful as SSR markers, provided that a sufficient number of polymorphic transposon insertion sites are identified in the species under investigation.

Transposon Display

Transposon Display, like AFLP, involves restriction digestion of the template DNA, ligation of adaptors, and subsequent PCR amplification. However, it differs from S-SAP by usually including an additional step which increases the frequency of restriction fragments related to transposon insertion. It can be carried out by using:

1) a rare restriction site present inside the transposon, ligation of a biotinylated adaptor and subsequent streptavidin capture of restriction fragments prior to PCR (Van den Broeck et al., 1998; Yephermov and Saedler, 2000);
2) a biotinylated transposon-specific primer in the first round of PCR, and subsequent streptavidin capture of amplified fragments (Yephermov and Saedler, 2000);
3) two rounds of PCR with nested transposon-specific primers (Casa et al., 2000; Biedler et al., 2003; Park et al., 2003; Kwon et al., 2005).

The third approach is currently the one most widely used, because it is relatively simple like AFLP, and also highly reliable and informative. Unlike the other techniques, Transposon Display has not yet been used for LTR-retrotransposons. It has, though, been widely used for other groups of transposons, including:

- non-LTR retrotransposons such as *Cin4* in maize (Yephremov and Saedler, 2000);
- *Ac/Ds*-like transposons such as *dTph1* in the petunia (Van den Broeck et al., 1998; De Keukeleire et al., 2001);
- *En/Spm*-like transposons such as *Rim2/Hipa* in rice (Kwon et al., 2005); and
- *MITE*-like transposons such as:
  - *Heartbreaker* in maize (Casa et al., 2000, 2002);
  - *Pegasus* in the mosquito (Biedler et al., 2003); and
  - *Pangrangja* in rice (Park et al., 2003).

CONCLUDING REMARKS

Repetitive sequences make up a large part of the genome, up to 80% in certain species such as maize. Dozens or even hundreds of copies of members of some transposon families can be present in a single genome. Transposons can serve as a very rich source of identifiable polymorphisms. They might even be more useful than the microsatellite regions which are currently in wide use.

Polymorphic transposon insertion sites can be identified only after the transposons residing in the investigated genome have been targeted and sequenced. Genomes of model organisms are being thoroughly characterized in terms of the presence of transposons, the number of copies, and the level of polymorphism. Strategies have been developed for identifying related groups of transposons in non-model species.

Simple PCR assays have been used to detect several *hAT*-like and *En/Spm*-like transposons in various plant genomes (De Keukeleire et al., 2004; Staginnus et al., 2001). In both assays, the internal fragment of the

transposon was amplified and sequenced using PCR. The degenerate primer pairs used hybridized with the most highly conserved regions of transposase, and were designed on the basis of known elements. These results cannot be used directly to develop a molecular marker system. The sequence of the transposon has to be extended first. Polymorphic insertion sites can be identified only after the terminal regions of the transposon are known.

A more complicated protocol has been used to identify plant *Ty1-copia*-like LTR retrotransposons. This method involves restriction digestion, adaptor ligation, primary PCR amplification with a biotinylated primer, streptavidin capture, and secondary PCR amplification (Pearce et al., 1999). The advantage was that the retrotransposon fragments sequenced contained a portion of the LTR which could be used directly to design primers for the S-SAP system.

Transposition events are responsible for *de novo* variation. They may play a major role in increasing variability, especially in vegetatively propagated crops. Transposon activity may account for 80% of all mutations (Capy et al., 1998). Many of the spontaneous mutations detected in fruit and ornamental crops may have been caused by transposon activity. Mutations caused by transposition very often revert to the wild type when the transposon later perfectly excises itself. Mutations caused by transposition can also give rise to new phenotypes when the transposon either imperfectly excises itself or transposes itself to nearby locations in the genome.

The current state of information on transposons in fruit crops is very limited. There are, however, two well documented cases of mutations caused by transposons:

1) Parthenocarpy in the apple was caused by the insertion of an LTR retrotransposon into the *MdP1* gene (Yao et al., 2001). This gene is a MADS-box transcription factor homologous to *PISTILLATA* in *Arabidopsis*.
2) A putative *SINE* transposon has been identified in the promoter region of the *ACS1* gene (Sunako et al., 1999). Apples from trees homozygous for the resulting allele produce less ethylene, and therefore have a longer shelf life.

High throughput analysis using molecular markers derived from transposons can be performed on sports of the same cultivar. This may help identify differences caused transposon activity, including differences responsible for changes in phenotype.

The first attempt at developing a transposon-based marker system in the apple used IMP markers generated with two *MITE*-like transposons, *Ars1* and *Ars2* (Hadonou et al., 2003). Very recently, a report was published on the application of the S-SAP technique for characterisation and fingerprinting of apple clones obtained from ‘Gala’ and Braeburn’ cultivars (Venturi et al., 2006). Application of the retrotransposon-based molecular marker system enabled identification of polymorphisms differentiating clones showing altered fruit color, while previous attempts towards using other marker

systems, i.e. RFLP, RAPD, SSR, and AFLP, for the same purpose were unsuccessful.

More information on the sequences of the genomes of economically important species including fruit and ornamental crops will facilitate the development of new systems for detecting the insertion of transposons belonging to different families.

Acknowledgement: The author wishes to thank Prof. Maria Klein for critical reading and helpful suggestions on the manuscript.

REFERENCES

Biedler J., Qi Y., Holligan D., della Torre A., Wessler S., Tu Z. 2003. Transposable element (TE) display and rapid detection of TE insertion polymorphism in the *Anopheles gambiae* species complex. INSECT MOL. BIOL. 12: 211-216.

Capy P., Bazin C., Higuet D., Langin T. 1998. Dynamics and Evolution of Transposable Elements. Chapman and Hall, New York, 197p.

Casa A.M., Brouwer C., Nagel A., Wang L., Zhang Q., Kresovich S., Wessler S.R. 2000. The MITE family *Heartbreaker* (Hbr): Molecular markers in Maize. PROC. NATL. ACAD. SCI. USA 97: 10083-10089.

Casa A.M., Mitchell S.E., Smith O.S., Register III J.C., Wessler S.R., Kresovich S. 2002. Evaluation of *Hbr* (MITE) markers for assessment of genetic relationships among maize (*Zea mays* L.) inbred lines. THEOR. APPL. GENET. 104: 104-110.

Chang R.Y., O' Donoughue L.S., Bureau T.E. 2001. Inter-MITE polymorphisms (IMP): a high throughput transposon-based genome mapping and fingerprinting approach. THEOR. APPL. GENET. 102: 773-781.

De Keukeleire P., De Schepper S., Gielis J., Gerats T. 2004. A PCR-based assay to detect *hAT*-like transposon sequences in plants. CHROMOSOME RES. 12: 117-123.

De Keukeleire P., Maes T., Sauer M., Zethof J., Van Montagu M., Gerats T. 2001. Analysis by Transposon Display of the behavior of the *dTph1* element family during ontogeny and inbreeding of *Petunia hybrida*. MOL. GEN. GENOMICS 265: 72-81.

Flavell A.J., Knox M.R., Pearce S.R., Ellis T.H.N. 1998. Retrotransposon-based insertion polymorphisms (RBIP) for high throughput marker analysis. PLANT J. 16: 643-650.

Hadonou A.M., Gittins J.R., Hiles E.R., James D.J. 2003. Two apple repetitive sequence elements: characterisation and potential use as genetic markers. EUPHYTICA 131: 177-187.

Jiang N., Bao Z., Zhang X., Hirochika H., Eddy S.R., Mc Couch S.R., Wessler S.R. 2003. An active DNA transposon family in rice. NATURE 421: 163-167.

Kalendar R., Grob T., Regina M., Suoniemi A., Schulman A. 1999. IRAP and REMAP: two new retrotransposon-based DNA fingerprinting techniques. THEOR. APPL. GENET. 98: 704-711.

Kikuchi K., Terauchi K., Wada M., Hirano H.Y. 2003. The plant MITE *mPING* is mobilized in anther culture. NATURE 421: 167-170.

Kwon S.J., Park K.C., Kim J.H, Lee J.K., Kim N.S. 2005. *Rim2/Hipa* CACTA transposon display; A new genetic marker technique in *Oryza* species. BMC Genetics 6: 15 doi:10.1186/1471-2156-6-15.

Leigh F., Kalendar R., Lea V., Lee D., Donini P., Schulman A.H. 2003. Comparison of the utility of barley retrotransposon families for genetic analysis by molecular marker techniques. MOL. GEN. GENOMICS 269: 464-474.

Manninen O., Kalendar R., Robinson J., Schulman A.H. 2000. Application of *BARE-1* retrotransposon markers to the mapping of a major resistance gene for net blotch in barley. MOL. GEN. GENET. 264: 325-334.

Mc Clintock B. 1948. Mutable loci in maize. CARNEGIE INST. WASH. YEARBOOK 47: 155-169.

Orgel L.E., Crick F.H.C. 1980. Selfish DNA: the ultimate parasite. NATURE 284: 604-607.

Park K.C., Kim N.H., Cho Y.S., Kang K.H., Lee J.K., Kim N.S. 2003. Genetic variations of AA genome *Oryza* species measured by MITE-AFLP. THEOR. APPL. GENET. 107: 203-209.

Pearce S.R., Knox M., Ellis T.H. N., Flavell A.J., Kumar A. 2000. Pea *Ty1-copia* group of retrotransposons: transpositional activity and use as markers to study genetic diversity in *Pisum*. MOL. GEN. GENET. 263: 898-907.

Pearce S.R., Stuart- Rogers, C., Knox M.R., Kumar A., Ellis T.H.N., Flavell A. 1999. Rapid isolation of plant *Ty1-copia* group retrotransposon elements for molecular marker studies. PLANT J. 19: 711-717.

Porceddu A., Albertini E., Barcaccia G., Marconi G., Bertoli F.B., Veronesi F. 2002. Development of S-SAP markers based on an LTR-like sequence from *Medicago sativa* L. MOL. GEN. GENOMICS 267: 107-114.

Queen R.A., Gribbon B.M., James C., Jack P., Flavell A.J. 2004. Retrotransposon-based molecular markers for linkage and genetic diversity analysis in wheat. MOL. GEN. GENOMICS 271: 91-97.

Staginnus C., Huettel B., Desel C., Schmidt T., Kahl G. 2001. A PCR-based assay to detect *En/Spm*-like transposon sequences in plants. CHROMOSOME RES. 9: 591-605.

Sunako T., Sakuraba W., Senda M., Akada S., Ishikawa R., Niizeki M., Harada T. 1999. An allele of the ripening specific 1-aminocyclopropane-1-carboxylic acid synthase gene (*ACS1*) in apple fruit with long storage life. PLANT PHYSIOL. 119: 1297-1303.

Syed N.H., Sureshundar S., Wilkinson M.J., Bhau B.S., Cavalcanti J.J.V., Flavell A.J. 2005. *Ty1-copia* retrotransposon-based SSAP marker development in cashew (*Anacardium occidentale* L.). THEOR. APPL. GENET. 110: 1195-1202.

Van der Broeck D., Maes T., Sauer M., Zethoff J., De Keukeleire P., D' Hauw M., Van Montagu M., Gerats T. 1998. Transposon Display identifies individual transposable elements in high copy number lines. PLANT J. 13: 121-129.

Venturi S., Dondini L., Donini P., Sansavini S. 2006. Retrotransposon characterisation and fingerprinting of apple clones by S-SAP markers. THEOR. APPL. GENET. 112: 440-444.

Vos P., Hogers R., Bleeker M., Reijans M., van de Lee T., Hornes A., Frijters J., Pot J., Peleman J., Kuiper M., Zabeau M. 1995. AFLP: a new technique for DNA fingerprinting. NUCL. ACIDS RES. 23: 4407-4414.

Waugh R., McLean K., Flavell A. J., Pearce S.R., Kumar A., Thomas B.T., Powell W. 1997. Genetic distribution of *BARE-1* retrotransposable elements in the barley

genome revealed by sequence-specific amplification polymorphisms (S-SAP). MOL. GEN. GENOMICS 253: 687-694.

Wessler S.R., Bureau T.E., White S.E. 1995. LTR-retrotransposons and MITEs: important players in the evolution of plant genomes. CURR. OPIN. GENET. DEV. 5: 814-21.

Yao J.L., Dong Y.H., Morris B.A.M. 2001. Parthenocarpic apple fruit production conferred by transposon insertion mutations in a MADS-box transcription factor. PROC. NATL. ACAD. SCI. USA 98: 1306-1311.

Yephremov A., Saedler H. 2000. Display and isolation of transposon-flanking sequences starting from genomic DNA or RNA. PLANT J. 21: 658-672.

POLIMORFIZM TRANSPOZONÓW JAKO NOWE ŹRÓDŁO MARKERÓW MOLEKULARNYCH

Dariusz Grzebelus

STRESZCZENIE

Transpozony będące niewielkimi, ruchomymi elementami genomu roślinnego mają duże znaczenie dla ewolucji genomów. Każda transpozycja (przeniesienie fragmentu DNA) może być źródłem zmienności w świecie roślin. W pracy scharakteryzowane są klasy transpozonów (retranspozony i transpozony DNA), mechanizmy ich przenoszenia oraz sposoby wykrywania obecności transpozonów w genomie roślinnym. Identyfikacja transpozonów oraz generowanie sprzężonych z nimi markerów molekularnych umożliwia nie tylko badanie bioróżnorodności, ale także tworzenie map genetycznych roślin różnych gatunków.

Słowa kluczowe: elementy ruchome genomu, bioróżnorodność, mapowanie genomu